



University of  
Zurich<sup>UZH</sup>

English Department

---

# The genitive's choice

Exploring *of*-genitive and Saxon genitive

Hans Martin Lehmann, Gerold Schneider, Melanie Röthlisberger

ICAME 2012, Leuven

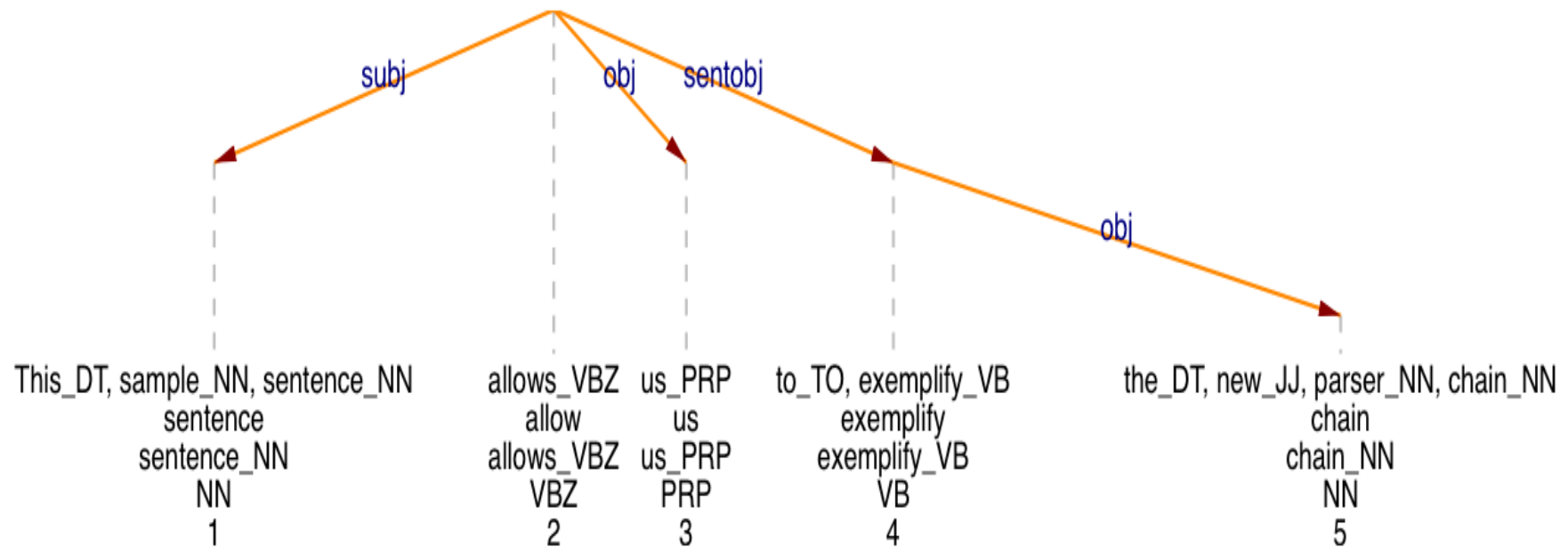
# Data

source	abbreviation	word-count
<b>CNN Transcripts</b>	cnn0311	530,116,720
<b>Guardian</b>	guardian	266,005,393
<b>New York Times</b>	nyt	254,797,266
<b>BNC XML written</b>	bncxwri	92,750,303
<b>LA Times</b>	latm	44,466,656
<b>The Boston Globe</b>	bogl	39,489,656
<b>[London] Times</b>	tlnd	38,360,679
<b>The Daily Mail</b>	tdma	23,177,008
<b>USA Today</b>	usat	19,100,759
<b>Sunday Times</b>	tlns	18,519,652
<b>The Mail on Sunday</b>	tmos	6,612,108
<b>total</b>		<b>1,333,396,200</b>

# Annotation: Tagging, Lemmatization, Chunking

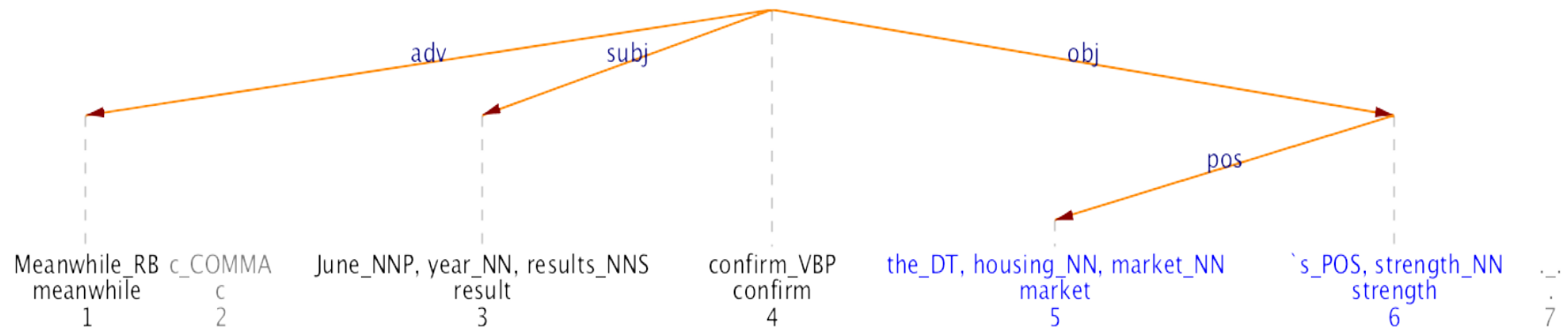
```
<?xml version="1.0"?>
<text>
  <p>
    <s id="s1">
      <ng>
        <w pws="yes" id="w1" p="DT">This</w>
        <w l="sample" pws="yes" id="w6" p="NN" headn="yes">sample</w>
        <w l="sentence" pws="yes" id="w13" p="NN" headn="yes">sentence</w>
      </ng>
      <vg tense="pres" voice="act" asp="simple" modal="no">
        <w l="allow" pws="yes" id="w22" p="VBZ" headv="yes">allows</w>
      </vg>
      <ng>
        <w pws="yes" id="w29" p="PRP">us</w>
      </ng>
      <vg tense="inf" voice="act" asp="simple" modal="no">
        <w pws="yes" id="w32" p="TO">to</w>
        <w l="exemplify" pws="yes" id="w35" p="VB" headv="yes">exemplify</w>
      </vg>
      <ng>
        <w pws="yes" id="w45" p="DT">the</w>
        <w pws="yes" id="w49" p="JJ">new</w>
        <w l="parser" pws="yes" id="w53" p="NN" headn="yes">parser</w>
        <w l="chain" pws="yes" id="w60" p="NN" headn="yes">chain</w>
      </ng>
      <w sb="true" pws="no" id="w65" p=".">.</w>
    </s>
  </p>
</text>
```

# Annotation: Parsing



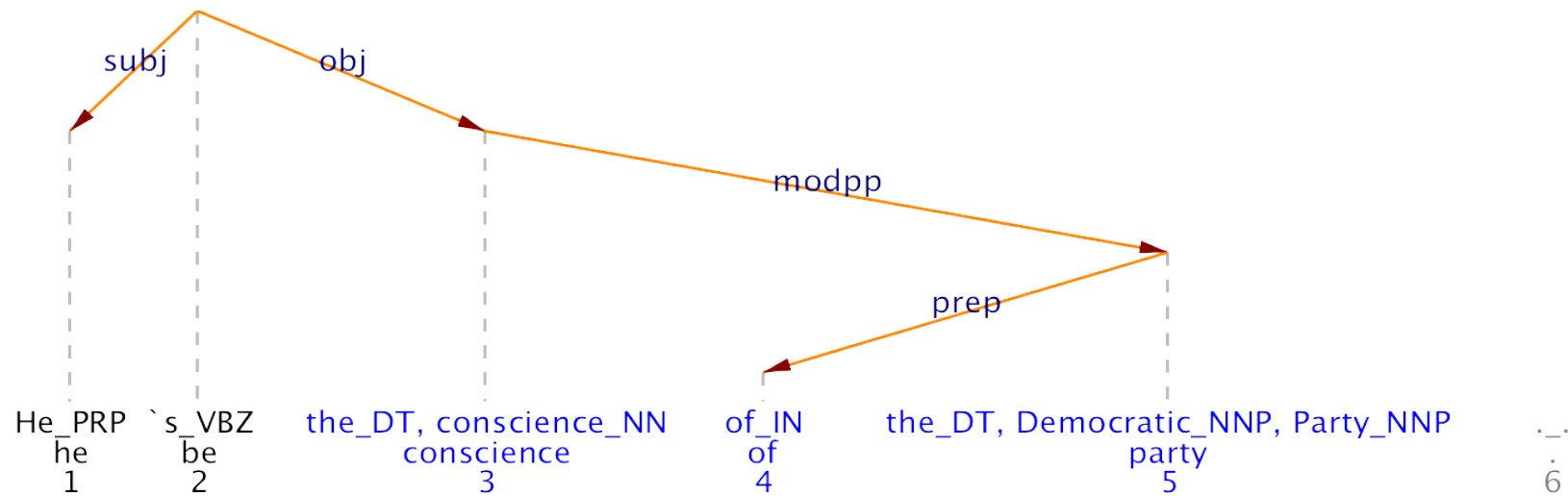
<b>words</b>	1333396200
<b>s-units</b>	76305807
<b>chunks</b>	1029305164
<b>relations</b>	1223855948
<b>rels ce</b>	851850096
<b>rels ci</b>	372005852

# Retrieval: Saxon genitive (26'078'946 instances)



- 2) Meanwhile, June year results confirm the housing market's strength.  
(tdma:SP:15362907)

## Retrieval: of-genitive (6'771'816 instances)



1) He's the conscience of the Democratic Party. (cnn0311::46866921)

# Evaluation of random sample

Recall & precision of saxon and of-pp.

Alternation yes/no.

X % of saxons can be expressed as normans

Y % of normans can be expressed as saxons

Inter annotator (dis-) agreement.

# Automatic Annotation / Over-coding

sID	28455151	corpus	cnn0311	reference	
-----	----------	--------	---------	-----------	--

stext	There is an effort now under way in Washington, led by the securities industry, to eliminate the role of state regulators.				
-------	--	--	--	--	--

prevnext

---

possorDET	possorMOD	possor		
na_no	state	regulator		
posseePrep	na_no			
possorID	28455151#	possorLen	24	

---

posseeDET	posseeMOD	possee		
the	na_no	role		
posseeModPP	of			
posseeID	28455151#	posseeLen	15	

---

govRel	obj	region	us	
govHead	eliminate	spowri	spoken	
govHeadID	28455151#14	dom	news	
govHeadPrep	na_no			



# Lexical Types

# Genitives with the preposition *for* and *by*

# Complex Genitives

## From the types back to the tokens

	of-gen	saxon	total
all	26078946	6771816	32850762
cut_off a	4034372	1922162	5956534
cut_off b	598668	582102	1180770

cut off a:

$f_{\text{sax}} > 0$  and  $f_{\text{of}} > 0$

$f_{\text{of}} > f_{\text{by}}$  and  $f_{\text{of}} > f_{\text{for}}$

possee: exclusion of core sortals

complex possessives are excluded

# Coding tokens for logistic regression

## Results: Cutoff A

## Results: Cutoff B

# Chapter Title



# Table of Contents

- dummy text
- dummy text
- dummy text
- dummy text
- dummy text
- dummy text
- dummy text
- dummy text

## The title of this slide can take up two lines

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce varius arcu sit amet nisl iaculis consectetur. Aenean iaculis, diam gravida accumsan facilisis, dolor purus rutrum enim, non egestas sem metus vitae felis. Proin consectetur augue et enim viverra sagittis. Aliquam ac nisi tellus.

Fusce dapibus turpis nisi, vitae dapibus velit. Cras sit amet consequat nulla. Nam dui lectus, adipiscing id volutpat id, luctus in metus. Nunc a ante ut est viverra sagittis.

Proin ullamcorper, massa eu dictum viverra, lacus elit bibendum ligula, a cursus libero ligula ac dui. Praesent leo eros, lacinia at convallis nec, egestas at dolor. Mauris quis sapien sem, at ultrices erat. Proin ultricies, diam vel aliquam posuere, ipsum eros porttitor est, eget condimentum massa enim eu libero.

[www.uzh.ch](http://www.uzh.ch)